CHƯƠNG 2:
MÔ HÌNH HỒI QUY BỘI

GV: Trương Bích Phương

1. Mô hình hồi quy 3 biến

1.1. Hàm hồi quy tổng thể (PRF)

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + U_i$$

Trong đó

- Y là biến phụ thuộc
- X_2,X_3 là các biến độc lập
- X_{2i}, X_{3i} là giá trị thực tế của X_2, X_3
- U_i là các sai số ngẫu nhiên

1.2. Các giả thiết của mô hình

- Giá trị trung bình của đại lượng ngẫu nhiên U_i bằng 0
- Phương sai của U_i không thay đổi
- Không có sự tương quan giữa các U_i
- Không có sự tương quan (cộng tuyến) giữa X_2 và X_3
- Không có sự tương quan giữa các U_i và X_2,X_3

1.3. Ước lượng các tham số

Chúng ta sử dụng phương pháp bình phương nhỏ nhất OLS

$$PRF: Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + U_i$$

Hàm hồi quy mẫu tương ứng sẽ là :

$$SRF: Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + e_i$$

Hay: $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i}$

$$e_i = Y_i - \hat{Y}_i = Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i}$$

Theo phương pháp OLS thì các tham số $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$ được chọn sao cho

$$\sum e_i^2 = \sum \left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i}\right)^2 \rightarrow \min$$

Như vậy , công thức tính của các tham số như sau :

Ký hiệu: $y_i = Y_i - \overline{Y}$

$$x_{2i} = X_{2i} - \overline{X}_2$$

$$x_{3i} = X_{3i} - \overline{X}_3$$

$\hat{\beta}_2 =$

$\hat{\beta}_3 =$

$\hat{\beta}_1 =$

Người ta chứng minh được

$$\sum x_{2i}^2 = \sum X_{2i}^2 - n\left(\overline{X}_2\right)^2$$

$$\sum x_{3i}^2 = \sum X_{3i}^2 - n\left(\overline{X}_3\right)^2$$

$$\sum y_i^2 = \sum Y_i^2 - n\left(\overline{Y}\right)^2$$

$$\sum x_{2i}x_{3i} = \sum X_{2i}X_{3i} - n\overline{X}_2\overline{X}_3$$

$$\sum y_i x_{2i} = \sum Y_i X_{2i} - n\overline{YX_2}$$

$$\sum y_i x_{3i} = \sum Y_i X_{3i} - n\overline{YX_3}$$

Ví dụ

Bảng dưới đây cho các số liệu về doanh số bán (Y), chi phí chào hàng (X_2) và chi phí quảng cáo (X_3) của một công ty

Hãy ước lượng hàm hồi quy tuyến tính của doanh số bán theo chi phí chào hàng và chi phí quảng cáo

Doanh số bán Y_i (trđ)	Chi phí chào hàng X_2	Chi phí quảng cáo X_3
1270	100	180
1490	106	248
1060	60	190
1626	160	240
1020	70	150
1800	170	260
1610	140	250
1280	120	160
1390	116	170
1440	120	230
1590	140	220
1380	150	150

Từ số liệu trên, ta tính được các tổng như sau :

1.4. Hệ số xác định của mô hình

$TSS =$

$ESS =$

$RSS =$

$R^2 =$

1.4. Hệ số xác định của mô hình

Đối với mô hình hồi quy bội, người ta tính R^2 có hiệu chỉnh như sau :

$$\overline{R}^2 =$$

k là số tham số trong mô hình

Khi *k>1* thì:

1.4. Hệ số xác định của mô hình

Ví dụ : Tính hệ số xác định của mô hình hồi quy theo số liệu của ví dụ trước

1.4. Hệ số xác định của mô hình

1.5. Phương sai của hệ số hồi quy

Phương sai của các tham số hồi quy được tính theo các công thức sau:

$$\hat{\sigma}^2_{\hat{\beta}_1} = \hat{\sigma}^2 \left[\frac{1}{n} + \frac{\overline{X}_2^2 \sum x_{3i}^2 + \overline{X}_3^2 \sum x_{2i}^2 - 2\overline{X}_2 \overline{X}_3 \sum x_{2i} x_{3i}}{\sum x_{2i}^2 \sum x_{3i}^2 - \left(\sum x_{2i} x_{3i} \right)^2} \right]$$

$$se(\hat{\beta}_1) = \sqrt{\hat{\sigma}^2_{\hat{\beta}_1}}$$

Với $\hat{\sigma}^2 = \dfrac{RSS}{n-3}$

1.5. Phương sai của hệ số hồi quy

$$\hat{\sigma}^2_{\hat{\beta}_2} = \hat{\sigma}^2 \left[\frac{\sum x_{3i}^2}{\sum x_{2i}^2 \sum x_{3i}^2 - \left(\sum x_{2i} x_{3i}\right)^2} \right]$$

$$se(\hat{\beta}_2) = \sqrt{\hat{\sigma}^2_{\hat{\beta}_2}}$$

Với $\hat{\sigma}^2 = \dfrac{RSS}{n-3}$

1.5. Phương sai của hệ số hồi quy

$$\hat{\sigma}^2_{\hat{\beta}_3} = \hat{\sigma}^2 \left[\frac{\sum x_{2i}^2}{\sum x_{2i}^2 \sum x_{3i}^2 - \left(\sum x_{2i} x_{3i}\right)^2} \right]$$

$$se(\hat{\beta}_3) = \sqrt{\hat{\sigma}^2_{\hat{\beta}_3}}$$

Với $\hat{\sigma}^2 = \dfrac{RSS}{n-3}$

1.6. Khoảng tin cậy của các hệ số hồi quy

Khoảng tin cậy của β_1 Với độ tin cậy là 1-α

$$\left(\hat{\beta}_1 - t_{\frac{\alpha}{2}} \times se(\hat{\beta}_1); \hat{\beta}_1 + t_{\frac{\alpha}{2}} \times se(\hat{\beta}_1)\right)$$

Khoảng tin cậy của β_2

$$\left(\hat{\beta}_2 - t_{\frac{\alpha}{2}} \times se(\hat{\beta}_2); \hat{\beta}_2 + t_{\frac{\alpha}{2}} \times se(\hat{\beta}_2)\right)$$

1.6. Khoảng tin cậy của các hệ số hồi quy

Khoảng tin cậy của β_3

$$\left(\hat{\beta}_3 - t_{\frac{\alpha}{2}} \times se(\hat{\beta}_3); \hat{\beta}_3 + t_{\frac{\alpha}{2}} \times se(\hat{\beta}_3) \right)$$

Lưu ý khi tra bảng T-Student, trong trường hợp hàm hồi quy 3 biến thì bậc tự do là **(n-3)**

1.6. Khoảng tin cậy của các hệ số hồi quy

Ví dụ : Tính khoảng tin cậy của β_2 và β_3 mô hình hồi quy theo số liệu của ví dụ trước với độ tin cậy 95%

1.6. Khoảng tin cậy của các hệ số hồi quy

1.7. Kiểm định giả thiết

a) Kiểm định giả thiết về $\beta_1, \beta_2\ \beta_3$

$H_o: \beta_i = \beta_o$

$H_1: \beta_i \neq \beta_o$

Bước 1 : Lập khoảng tin cậy

Bước 2 : Nếu β_0 thuộc khoảng tin cậy thì chấp nhận H_o. Nếu β_0 không thuộc khoảng tin cậy thì bác bỏ H_o

Ví dụ : (theo số liệu trước), yêu cầu kiểm định các giả thiết β_2 , β_3 Với độ tin cậy 95%

1.7. Kiểm định giả thiết

b) <u>Kiểm định giả thiết về R^2</u>

$H_o: R^2 = 0$

$H_1: R^2 \neq 0$

<u>Bước 1</u> : tính

<u>Bước 2</u> : Tra bảng tìm F(2,n-3), mức ý nghĩa là α

<u>Bước 3</u> : Nếu $F > F(2,n-3)$, bác bỏ H_0

Nếu $F \leq F(2,n-3)$, chấp nhận H_0

1.7. Kiểm định giả thiết

b) Kiểm định giả thiết về R^2

Ví dụ: Yêu cầu kiểm định giả thiết

$H_o: R^2 = 0$
$H_1: R^2 \neq 0$

Độ tin cậy là 95%

2. Một số dạng hàm

2.1. Hàm sản xuất Cobb-Douglas

Hàm sản xuất Cobb-Douglas được biểu diễn như sau:

$$Y_i = \beta_1 X_{2i}^{\beta_2} X_{3i}^{\beta_3} e^{U_i}$$

Trong đó : Y_i : sản lượng của doanh nghiệp

X_{2i} : lượng vốn

X_{3i} : lượng lao động

U_i : sai số ngẫu nhiên

Hàm sản xuất Cobb-Douglas có thể đưa được về dạng tuyến tính bằng cách lấy logarit hai vế

2. Một số dạng hàm

2.1. Hàm sản xuất Cobb-Douglas

$$\ln Y_i = \ln \beta_1 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + U_i$$

Đặt $Y_i^* = \ln Y_i$

$$\beta_1^* = \ln \beta_1$$

$$X_{2i}^* = \ln X_{2i}$$

$$X_{3i}^* = \ln X_{3i}$$

Dạng tuyến tính sẽ là :

$$Y_i^* = \beta_1^* + \beta_2 X_{2i}^* + \beta_3 X_{3i}^* + U_i$$

2.2. Hàm hồi quy đa thức bậc 2

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + U_i$$

Mặc dù chỉ có một biến độc lập X_i nhưng nó xuất hiện với các luỹ thừa khác nhau khiến cho mô hình trở thành hồi quy ba biến

3. HỐI QUY TUYẾN TÍNH K BIẾN

3.1. Hàm hồi quy tổng thể (PRF)

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + ... + \beta_k X_{ki} + U_i$$

Trong đó

- Y là biến phụ thuộc
- $X_2, X_{3,...,} X_k$ là các biến độc lập
- U_i là các sai số ngẫu nhiên
- β_1 :Hệ số tự do

$\beta_2, \beta_3, ..., \beta_k$ là các hệ số hồi quy riêng

3.1. Hàm hồi quy tổng thể (PRF)

Ký hiệu

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_n \end{pmatrix} \qquad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_k \end{pmatrix} \quad U = \begin{pmatrix} U_1 \\ U_2 \\ \dots \\ U_n \end{pmatrix}$$

3.1. Hàm hồi quy tổng thể (PRF)

Và

$$X = \begin{pmatrix} 1 & X_{21} & X_{31} & \cdots & X_{k1} \\ 1 & X_{22} & X_{32} & \cdots & X_{k2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & X_{2n} & X_{3n} & \cdots & X_{kn} \end{pmatrix}$$

3.1. Hàm hồi quy tổng thể (PRF)

Khi đó , hệ thống các quan sát có thể được viết lại dưới dạng :

$$Y = X.\beta + U$$

3.2. Các giả thiết của mô hình hồi quy k biến

Giả thiết 1 : Các biến độc lập X_1, X_2,…,X_k đã cho và không ngẫu nhiên

Giả thiết 2 : Các sai số ngẫu nhiên U_i có giá trị trung bình bằng 0 và có phương sai không đổi

Giả thiết 3: Không có sự tương quan giữa các sai số U_i

Giả thiết 4 : Không có hiện tượng cộng tuyến giữa các biến độc lập X_2, X_3,…,X_k

Giả thiết 5 : Không có tương quan giữa các biến độc lập X_2,X_3,…,X_k với các sai số ngẫu nhiên U_i

3.2. Các giả thiết của mô hình hồi quy k biến

Hàm hồi quy mẫu :

SRF: $$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \ldots + \hat{\beta}_k X_{ki} + e_i$$

hoặc: $$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \ldots + \hat{\beta}_k X_{ki}$$

Hay : (Viết dưới dạng ma trận)

$$Y = X\hat{\beta} + e$$

3.3. Ước lượng các tham số

Với

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \dots \\ \hat{\beta}_k \end{pmatrix} \qquad e = \begin{pmatrix} e_1 \\ e_2 \\ \dots \\ e_n \end{pmatrix}$$

3.3. Ước lượng các tham số

SRF: $Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + ... + \hat{\beta}_k X_{ki} + e_i$

hoặc: $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + ... + \hat{\beta}_k X_{ki}$

Khi đó

$$e_i = (Y_i - \hat{Y}_i)$$

$$= Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i} - ... - \hat{\beta}_k X_{ki}$$

3.3. Ước lượng các tham số

Theo phương pháp OLS thì các tham số

$\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \ldots, \hat{\beta}_k$ được chọn sao cho

$$\sum e_i^2 = \sum \left(Y_i - \hat{Y}_i \right)^2$$

$$= \sum \left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i} - \ldots - \hat{\beta}_k X_{ki} \right)^2$$

$$\to \min$$

Ta ký hiệu $X^T, Y^T, \hat{\beta}^T, e^T$ là các ma trận chuyển vị của $X, Y, \hat{\beta}, e$

Tức là

$$Y^T = (Y_1, Y_2, \ldots, Y_n)$$

$$e^T = (e_1, e_2, \ldots, e_n)$$

$$\hat{\beta}^T = (\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k)$$

3.3. Ước lượng các tham số

$$X = \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ X_{21} & X_{22} & X_{23} & \dots & X_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ X_{k1} & X_{k2} & X_{k3} & \dots & X_{kn} \end{pmatrix}$$

Khi đó :

$$\hat{\beta} = (\boldsymbol{X}^{T}\boldsymbol{X})^{-1}\boldsymbol{X}^{T}\boldsymbol{Y}$$

Trong đó $(X^T X)$ là ma trận có dạng

$$X = \begin{pmatrix} n & \sum X_{2i} & \sum X_{3i} & \dots & \sum X_{ki} \\ \sum X_{2i} & \sum X_{2i}^2 & \sum X_{2i} X_{3i} & \dots & \sum X_{2i} X_{ki} \\ \dots & \dots & \dots & \dots & \dots \\ \sum X_{ki} & \sum X_{ki} X_{2i} & \sum X_{ki} X_{3i} & \dots & \sum X_{ki}^2 \end{pmatrix}$$

3.3. Ước lượng các tham số

Ví dụ minh hoạ

Bảng dưới đây cho các số liệu về lượng hàng bán được của một loại hàng hóa(Y), thu nhập của người tiêu dùng (X_2) và giá bán của loại hàng này (X_3)

Tìm hàm hồi quy tuyến tính

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i}$$

Y_i (tấn/tháng)	X_2 (triệu đồng/năm)	X_3(ngàn đồng/kg)
20	8	2
18	7	3
19	8	4
18	8	4
17	6	5
17	6	5
16	5	6
15	5	7
13	4	8
12	3	8

Giải Từ số liệu trên, ta tính được các tổng như sau:

3.4. Hệ số xác định của mô hình

$$TSS = Y^T Y - n(\overline{Y})^2$$

$$ESS = \hat{\beta}^T X^T Y - n(\overline{Y})^2$$

$$RSS =$$

$$R^2 =$$

3.5. Khoảng tin cậy và kiểm định giả thiết

Gọi c_{jj} là phần tử nằm ở dòng j cột j của ma trận $(X^TX)^{-1}$

Khi đó : $$\sigma^2_{\hat{\beta}_j} = \sigma^2.c_{jj} \approx \hat{\sigma}^2.c_{jj}$$

$$se(\hat{\beta}_j) = \sqrt{\sigma^2_{\hat{\beta}_j}}$$

Với $\hat{\sigma}^2 = \dfrac{RSS}{n-k}$ (k là số tham số)

3.5. Khoảng tin cậy và kiểm định giả thiết

Khoảng tin cậy của β_j là

$$(\hat{\beta}_j - t_{\frac{\alpha}{2}} se(\hat{\beta}_j); \hat{\beta}_j + t_{\frac{\alpha}{2}} se(\hat{\beta}_j))$$

Hoặc tính giá trị tới hạn của β_j là

$$t = \frac{\hat{\beta}_j - \beta_j^*}{se(\hat{\beta}_j)}$$

Bậc tự do là (n-k)

3.5. Khoảng tin cậy và kiểm định giả thiết

Kiểm định giả thiết về R^2 $H_o: R^2 = 0$

$H_1: R^2 \neq 0$

Với độ tin cậy 1-α

Bước 1 : tính $F = \dfrac{R^2(n-k)}{(k-1)\left(1-R^2\right)}$

Bước 2 : Tra bảng tìm F(k-1,n-k), mức ý nghĩa là α

Bước 3 : Nếu F>F(k-1,n-k) , bác bỏ H_0 Nếu F≤F(k-1,n-k) , chấp nhận H_0

3.6. Vấn đề dự báo

Cho $$X_o = \begin{pmatrix} 1 \\ X_2^0 \\ \dots \\ X_n^0 \end{pmatrix}$$

Yêu cầu dự báo giá trị Y_0 của Y

3.6. Vấn đề dự báo

Dự báo điểm :

$$\hat{Y}_0 = \hat{\beta}_0 + \hat{\beta}_2 X_2^0 + \hat{\beta}_3 X_3^0 + \ldots + \hat{\beta}_k X_k^0$$

Dự báo khoảng :

$$(\hat{Y}_0 - t_{\frac{\alpha}{2}} se(\hat{Y}_0); \hat{\beta}_j + t_{\frac{\alpha}{2}} se(\hat{Y}_0))$$

Bậc tự do là (n-k)

3.6. Vấn đề dự báo

$$\sigma^2_{\hat{Y}_0} = \hat{\sigma}^2 X_0^T (X^T X)^{-1} X^0$$

$$se(\hat{Y}_0) = \sqrt{\sigma^2_{\hat{Y}_0}}$$

Ví dụ (số liệu trước)

Ví dụ : Tính khoảng tin cậy của β_2 theo số liệu của ví dụ trước với độ tin cậy 95%

Ví dụ (số liệu trước)

yêu cầu kiểm định các giả thiết

$H_o: \beta_2 = 0$

$H_1: \beta_2 \neq 0$

Với độ tin cậy 95%

Ví dụ (số liệu trước)

Yêu cầu kiểm định các giả thiết

$$H_o: R^2 = 0$$
$$H_1: R^2 \neq 0$$

Với độ tin cậy 95%

Ví dụ (số liệu trước)

Yêu cầu dự báo giá trị của Y khi $X_2=9$ và $X_3=9$ với độ tin cậy 95%

Y: Doanh thu, X_2 Chi phí quảng cáo, X_3= Lương nhân viên

Y	X2	X3	Y	X2	X3
126	**17**	**11**	**160**	**23**	**15**
148	**23**	**14**	**127**	**15**	**11**
105	**18**	**9**	**138**	**16**	**12**
162	**22**	**16**	**143**	**21**	**14**
101	**14**	**9**	**158**	**22**	**15**
175	**24**	**17**	**137**	**13**	**13**

1. Viết phương trình hồi quy
2. Tìm R^2 và R^2 hiệu chỉnh
3. Tìm khoảng tin cậy của các hệ số hồi quy
4. Kiểm định giả thiết H_0 $\beta_1 = 0$, H_0 $\beta_2 = 0$ với α=5%
5. Dự báo doanh thu cty nếu lương 10tr, quảng cáo 100tr